An Unmanned Spacecraft Subsystem Cost Model for Advanced Mission Planning

G. Madrid, Jet Propulsion Laboratory, March 23, 1998

1. *Introduction*

As a NASA center, the Jet Propulsion Laboratory (JPL) is committed to the concept of developing and launching a continuously improving series of smaller robotic space exploration missions in shorter intervals of time (faster, better, cheaper). In order to plan and budget these advanced missions, JPL has begun an institutional initiative labeled "Develop New Products" (DNP). This institutional initiative involves an across the board paradigm shift in the process with which new projects are planned, designed, and implemented in an accelerated implementation cycle. A key factor in the planning of these missions is an accurate estimation of their cost so that an adequate, yet efficient, budget may be proposed that will not only be acceptable to NASA but will ensure a realistic implementation of a specific project within a predetermined project implementation schedule and risk envelope.

The project planning process has also been accelerated so that cost estimates may be produced within a one to two week cycle. This permits a second or third cost estimate to be produced that takes into account technology-cost trades vs. science objectives derived from the advanced planning deliberations in which the cost estimators play a key role. Once converged, this process leads to a budget estimate that has achieved a certain degree of consensus within the JPL community and its industrial partners prior to entering the proposal stage. Because of this, the probability of approval of the proposal is greatly increased.

The main instrument for carrying out this advanced planning process is a team of spacecraft and ground system engineering experts termed "Team X" at JPL.
The team members are key technical staff selected by the JPL technical divisions as having the expertise required to design and cost the subsystem to which they have been assigned. This team conducts its deliberations around a distributed workstation facility that interacts through a network in conjunction with a central data base and a documentarian. This arrangement permits the study leader and team members to interact in "real time" to develop a preliminary design and cost estimate within a week. Such a process would normally have taken from three to four months under the previous paradigm. A large portion of the proposals reviewed by Team X are of the DNP type. In order for a new project to be termed "DNP", the proposal must establish that the implementation (from concept to launch) can be accomplished within 33 months and the final cost estimate must fall into a cost range between $120M and $500M, not including the launch vehicle. Projects falling outside this range are processed using other more pertinent models.

are examined by the Study Lead and the Team X system engineer and may be overridden by them.

The cost estimation process uses differing approaches to predicting cost based on the portion of the work breakdown structure (WBS) being estimated. The basic methods used for estimating the cost of the distinct portions of the total project cost are:

1. Statistically-based algorithms from the previous Deep Space Cost Model that have been adjusted to conform to the DNP paradigm. This type of algorithm are termed historically-based algorithms (Hist. Based Algo.)
2. A non-statistical algorithm based on a quasi-grass-roots-based estimate and expert opinion formulated in consultation with technical specialists in the area of the project component being assessed. The algorithm is based on an evaluation of actual data and the design of the function being performed but which does not have sufficient structure to formulate a model at this time.
3. The current Instrument Cost Model developed by Keith Warfield
4. The current Subsystem Cost Model developed by Leigh Rosenberg
5. The actual price of the item being assessed, as in the case for launch vehicles, where the cost to the government is either predetermined by agreement with the vendors or is the listed price for the service.

The following lists the components of the advanced project cost estimation process and the method used:

	Project Cost Component	Cost Est. Method
1.	Project Management and Administration	Hist.-Based Algo.
2.	Science and Science Team Activities	Quasi-GR-Based Algo.
3.	Project and Mission Engineering	Hist.-Based Algo.
4.	Payload (Instruments)	Instr. Cost Model
5.	Spacecraft (System & Subsystem Costs)	
	5.1 System Level Mgmt & Engrg	Hist.-Based Algo.
	5.2 S/C Subsystem Costs	**S/S Cost Model**
6.	Assembly, Test, and Launch Operations	Quasi-GR-Based Algo.
7.	Mission Operations Development	Quasi-GR-Based Algo.
8.	Launch Vehicle	Current Price

The discussion in this paper concerns itself solely with the spacecraft subsystem costs, item 5.2. Mr. Rosenberg's paper will discuss the overall process (items 1-8) while Mr. Warfield's paper will deal with the instrument model used in item 4.

This paper describes the subsystem portion of the unmanned mission and spacecraft implementation cost model used in this interactive environment that is consistemt with the DNP assumptions. This mission and spacecraft subsystem cost model was developed by Mr. Leigh Rosenberg of JPL. An adjunct instrument model was developed by Mr. Keith Warfield, also of JPL. Companion papers are being submitted by Mr. Rosenberg, Mr. Warfield, and other cost team members that describe other aspects of the new cost estimation environment including the historical and evolutionary aspects. The focus of this paper will be on the design and structure of the subsystem cost model itself.

2. *Model Overview*

Because no unmanned space missions have yet been fully implemented using the new spacecraft development lifecycle paradigm shift, the cost model used is not based on a historical data base of previously implemented missions. Rather, the model is based on a data base of the prior estimates of proposed missions that have been developed using the Team X process and that have been certified as viable candidates for future mission proposals. As a result, the model described here acts as a predictor of Team X results and is currently used to validate the on-going estimates being developed with respect to a consistency with the DNP Process, past predictions, and previously proposed designs.

The focus of this paper is on a subsystem cost model that is based on data obtained from the Team X process, not on the process or estimates obtained by the team. Although the model is a predictor of the planning team results, it was nonetheless designed as if the parameters and cost data were obtained from an as-built design. An effort is under way to validate model estimates obtained using the new paradigm as soon as mission implementation costs are available from more recent missions that do business under the new paradigm.

The Cost Model is linked to the Team X system and subsystem workstations so that the technical parameters required by the model are passed to the cost workstation which updates the estimates of the cost for each subsystem as the deliberations are in progress. The model cost estimates are then used as a comparator to the costs being estimated by the team and are kept separate from the team deliberations so as not to bias the results. The Model cost estimates used in this manner are calculated using algorithms derived from the cost estimation relationships (CER's) derived from the statistical analysis performed on the data base of DNP projects mentioned above.

Some of the non-technical project/system infrastructure costs used during the Team X sessions are estimated by algorithms derived from historical costs for similar type projects (scaled to the DNP project time phase constraints). Since they are a function of total system, subsystem, and instrument costs, the algorithms permit a quick assessment of the infrastructure costs as the subsystem costs are being developed. At the end of the deliberations the predicted infrastructure costs

3. *Cost Model Data Base*

The Subsystem Cost Data Base is a collection of all of the system and subsystem technical parameters, subsystem masses, and associated cost estimates obtained as the result of Team X deliberations from October 1996 through October 1997. Of the nineteen proposed unmanned deep space projects whose estimates and parameters are contained in the data base, seventeen have been selected for application for the cost model. Other project cost estimates produced by Team X during the period the data base was constructed were excluded due to their unique characteristics which did not entirely fit into the DNP mold. The data base parameters are continuously undergoing some fine tuning as Team X review of the design, results in modification to the parameters.

Table 1, below, lists the cost portion of the data base by project. Due to the sensitive nature of the cost data regarding projects, these are only identified by a placeholder identification as P1, P2, etc.

Table 1. Subsystem Data Base Cost Summary

		Subsystem Costs (FY97$M)									
Project	Tot$M	ADCS	C&DH		Telcom	Power	Prop	Struct		Therm	Other
			Core	S/W				Core	MeB/U		
P1	91.4	17.8	12.7	2.0	15.0	6.7	15.7	12.3	3.4	5.8	
P2	96.7	17.7	9.1	1.4	13.0	14.6	19.6	9.5	3.4	8.4	
P3	95.0	13.4	4.4	2.0	13.9	15.2	20.7	10.3	3.4	2.2	9.5
P4	67.9			2.0		9.0	10.2				
P13	69.2	17.8	8.5	1.0	10.2	4.6	4.1	8.9	2.8	1.7	9.6
P14	54.8	11.9	2.4	0.8	10.4	5.5	10.2	8.3	3.2	2.1	
P15	33.4	6.1	2.1	0.6	5.0	5.3	3.5	7.4	1.7	1.7	
P16	51.7	10.2	2.9	0.8	8.1	6.1	9.7	9.4	2.8	1.7	
P17	36.7	6.2	2.4	0.7	5.4	5.8	3.5	8.2	2.8	1.7	
Avg	71.1	12.3	5.9	1.2	10.1	12.5	10.6	9.6	3.1	3.1	11.4
Std Dev	20.6	4.6	2.8	0.4	3.2	9.2	5.6	2.1	0.7	1.8	2.2
Max	98.2	22.3	12.7	2.0	15.0	42.4	20.7	13.5	5.0	8.4	15.0
Min	33.4	6.1	2.1	0.6	5.0	4.6	3.4	4.5	1.7	1.7	9.5

Table 2, lists the instances of the design parameters, $\{\xi\}$, which have been selected as having a causal relationship to cost for all projects in the data base.

Table 2. Subsystem Data Base Values for Technical Parameters by Project

	Technical Parameters by Project																	
	Stab. Type	Mission Life	Power Type	Prop Type	ISP	Total Power Rq'd	Total Power Therm.	Payload Power	PL Data Rate	DL Data Rate	Data Volume	Pnting Control	Pntng Knwldg	Band Type	Redun-dancy	Rad. Dose	S/C Mass	S/S Mass
Units	ordinal	Yrs	ordinal	ordinal	sec.	watts	watts	watts	kbps	kbps	Gb	arcsec	arcsec	ordinal	ordinal	krads	kg	kg
Param->	ξ_1	ξ_2	ξ_3	ξ_4	ξ_5	ξ_6	ξ_7	ξ_8	ξ_9	ξ_{10}	ξ_{11}	ξ_{12}	ξ_{13}	ξ_{143}	ξ_{15}	ξ_{16}	ξ_{17}	ξ_{18}
Project																		
P1	3	5.2		Chem	325	291	291	52.4		600	30	900	360	Ka	High	125	1641	(see
P2	3	3.4	GaAs	SEP, Chem	3000	6199	305	69.9		200	30	1800	360	X	High	150	1085	Table
P3	3	7.1			3500	711	192	2		1.2	1.44	360	180.0	X/Ka		15		3)
P11	3	6		SEP, Chem	3500	1704	324	20.8		0.9	10	133	60	Ka	High	12	919	
P12	3	4.6				117	117			0.1					Selected		354	
P13	3	5	SS	N_2H_4	220	189	189	33.5	4	125	10	30	5	X/Ka	Selected	72	530	
P14	Spin	0.8	GaAs, Li ion		325	215	214.9	10	2	2	0.006	900	180	S	Single	5	1004	
P15	Spin	0.5	GaAs, Li ion	N_2H_4	220	97	97.3	5	0.01	0.03	0.006	3600	900	S	Single	5	478	
P16	Spin	0.8	GaAs, Li ion	bi-prop	325	153	153.1	5	2	2	128	3600	900	S	Selected	5	826.8	
P17	Spin	0.5	GaAs, Li ion	N_2H_4	220	115	115	5		20		3600	900	S	Single	5	491	
Avg	3.0	4.5	n/a	n/a	941.9	650.4	171.0	23.3	13.1	63.2	6.8	757.9	244.7	n/a	n/a	398	989	
Std Dev	0.0	2.2	n/a	n/a	1254.4	1508.0	81.0	21.3	34.3	154.0	10.2	911.0	225.6	n/a	n/a	1029	778	
Max	3.0	9.0	n/a	n/a	3500.0	6199.3	324.0	69.9	110.0	600.0	30.0	3600.0	900.0	n/a	n/a	4000	3528	
Min	3.0	0.5	n/a	n/a	217.0	54.5	54.5	2.0	0.0	0.0	0.0	30.0	5.0	n/a	n/a	5	146	

The subsystem mass plays a role as a cost estimation parameter in some instances. Table 3. lists the subsystem mass data in the data base. When applicable to a particular regression fit, the subsystem mass is used as one of the technical parameters for the regression fit.

Table 3. Subsystem Data Base Values for Mass

	Mass Values by Subsystem for each Project (kg)												Total
Project	ADCS	C&DH	Telcom	Power	Prop	Struct.	Therm	Payload	Contin-gency	Propel-lant	LV Adapter	Other	S/C Mass
P1	25.7	14.6	30.3	27.4	118.7	173.6	47	17.6	136.5	1044.2	5.5		1641
P2	37.5	17.5	14.3	104.3	127	169.3	79	27.9	173	326.5	8.8		1085
P3	18.7	8	17.3	81.2	134	158.6	20.8	72.2	159	371.8	0	19.2	1061
		5.7			66.9		12	21.6	79.4		0		692
P11	24.6	3.7	15.5	69.8	141.7	112.7	18.5	31	71.1	430.5	0		919
P12	8.7	2.8	14.2	13.1	13.4	116.6	45	17	69.2	45.2	8.8		354
P13	15.9	10.4	13.1	10.4	8.7	71.6	4.4	180	46.3	4.8	14.6	150	530
P14	6.9	11	17.1	21.5	69.8	144.7	12	221.6	106.3	392.9	0		1004
P15	1.9	1	10.4	15.3	12.2	88.6	7.8	256.6	48.4	35.5	0		478
P16	7.1	1.6	22.7	15.3	53.8	125.2	13	275.1	80.6	232.2	0		827
P17	5.8	10.4	10	16.4	12.2	94.1	7.8	250	54	30.1	0		491
Avg	16.8	7.3	14.1	45.2	74.2	138.1	21.0	122.9	110.4	415.8	9.5	66.5	989
Std Dev	13.7	5.1	5.9	48.3	60.7	72.4	20.4	109.0	69.7	563.0	16.1	59.2	778
Max	48.1	17.5	30.3	195.1	220.1	336.6	79.0	356.4	272.4	2296.4	51.0	150.0	3528
Min	1.9	1.0	7.2	10.4	7.4	14.1	3.2	4.9	17.0	4.8	0.0	19.2	146

4. *Model Construction*

In order to predict subsystem costs from the data presented in the data base, a model that relates subsystem cost to the parameters, $\{\xi\}$, in table 2 is required. The approach taken was to define a regression model function that could be used for each of the subsystems to predict cost within the parameter data domain. The cost data and the parameters relevant to each subsystem would form the basis for a first order regression fit that would result in an equation that would then be used to predict costs for that subsystem within the predictive constraints imposed by the fit. The total subsystem costs would then be obtained by summing all of the subsystem cost estimates.

A generalized first order multivariate linear regression function [Draper and Smith,1966, § 5.1] was used throughout. Although some of the relationships are non-linear, they may be transformed to this linear form (ie., they are intrinsically linear). It was determined, through analysis and experimentation, that this approach would provide very acceptable fits for the data set currently in the data base. This type of function is traditionally expressed as follows:

$$\eta_i = \beta_0 + \Sigma\beta_j X_{ij} \qquad (j=1,k) \qquad (4.1)$$

where η_i is the dependent variable, X_{ij} are the independent variables, β_j are the undetermined coefficients of X_{ij} to be determined by means of the linear regression process, and β_0 is a constant (also to be determined). The index, i, refers to a particular instance where a measurement of η_i occurs for the specific subsystem for which the linear estimation is being made.

Assume that Y_i is the measurement of η_i such that,

$$Y_i - \eta_i = \varepsilon_i \qquad (4.2)$$

where ε_i is the measurement error and errors are assumed to be additive and satisfy the Gauss- Markov assumptions [Beck and Arnold,1977, § 5.1.3].

This being the case, we may then express the regression function (4.1) as:

$$Y_i = \beta_0 + \Sigma\beta_j X_{ij} + \varepsilon_i \qquad (j=1,k) \qquad (4.3)$$

In the particular application in question, the following interpretation will be given to the variables and coefficients:

Y_i The instance, i, of a cost measurement, Y, for the subsystem under assessment. Yi, is considered an estimate of the regression function, ηi, of the parameter values $(X_i)_j$ pertaining to the specific instance.

X_{ij} Instances of the technical parameters selected from the set $\{\xi\}$ that have a causal relation to the cost, Y. The selected parameters are ordered from j=1, k, in the equation (4.3). This ordinal specification may be different than that used in the global set of parameters $\{\xi\}$ since only the parameters influencing the cost are selected.

β_j The coefficients of the linear regression equation for the subsystem being assessed that are to be estimated by means of the regression process.

ε_i The measurement error, $Y_i - \eta_i$.

This form (4.3), is the regression model to be used in the discussion that follows. Other model approaches (including non-linear) were examined but did not produce significant improvements in fit for the particular set of data being evaluated.

The linear regression estimation process operates on two sets of data defined from the data base. These are: 1) an nx1 matrix of the cost instances, Y_i, for the subsystem being assessed , and 2) a corresponding nx(k+1) matrix of the instances of the technical parameters, X_{ij} selected as being causal for this subsystem.

$$\mathbf{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \bullet \\ \bullet \\ \bullet \\ \bullet \\ \bullet \\ \bullet \\ \bullet \\ \bullet \\ \bullet \\ Y_n \end{bmatrix} \quad (4.4) \qquad \mathbf{X} = \begin{bmatrix} 1 & X_{11} & X_{12} & \bullet\ \bullet\ \bullet\ \bullet & X_{1k} \\ 1 & X_{21} & X_{21} & & \\ 1 & X_{31} & X_{32} & X_{33} & \\ \bullet & & & & \\ \bullet & & & & \\ \bullet & & & & \\ \bullet & & & & \\ \bullet & & & & \\ \bullet & & & & \\ \bullet & & & & \\ \bullet & & & & \\ \bullet & & & & \\ 1 & Xn1 & \bullet\ \bullet\ \bullet\ \bullet & \bullet\ \bullet\ \bullet\ \bullet & X_{nk} \end{bmatrix} (4.5)$$

Using these data as input, the linear estimation process solves for estimates of β_j that minimize ε_i. These estimated coefficients are termed, b_j. In general, the results of the regression estimation is expressed with the predictive equation:

$$y_i = b_o + \Sigma b_j x_{ij} \qquad (j=1,k) \qquad (4.6)$$

where y_i is the predicted cost for the subject subsystem at any instance i , based on the estimated parameter coefficients, b_j, and the parameters, x_{ij}, specified for that subsystem at that instance. When using the predictive equation, care must be taken to ensure that the parameters selected fall within the domain of the data base parameters.

When the subsystem costs have been individually estimated, the total spacecraft system costs may be calculated by summing the subsystem results. Additional costs for system management, system engineering, spares, integration and test, and operations support need to be added to complete the cost estimate for the spacecraft. These costs and the costs associated with the project infrastructure itself will be dealt with in a folow-on paper.

The basic processes in construction of the model were as follows:

1) Validate the model data base to ensure that all of the information is appropriate and accurate,
2) In consultation with subsystem technologists, establish the initial set of parameters, X_{ij}, casually related to estimating the cost of each subsystem Y_i (eg, mass, power generation, radiation dosage, etc.). Ensure that these are appropriately and accurately represented in the data base.
3) Determine the general regression function to be used (as above),
4) Conduct an evaluation strategy using the regression strategy selected to determine the "best" parameters to leave in the fit. In this case a modified backward elimination process was performed to reduce the set of parameters, X_{ij} to be considered to those resulting in a validated "best fit" and and whose t statistics indicate validate the hypothesis that E(bj)=0, consistent with a maximization of the Coefficient of Multiple Determination, (R^2). Standard F- and t- test constraints for fit and coefficient validity were utilized.
5) Validate the resulting model against expected behavior within the valid range of the parameters. The model behavior is checked against independent subsystem estimates provided by the expert for that subsystem.
6) Reconstruct any of the model equations based on any new information obtained in the process of validating the model equation in (5).
7) The entire set of subsystem costs are then validated against the data base itself to ensure that the difference of the costs obtained vs. the data base costs for a particular project are within the expected variance of the model.

The current model equations will be updated as improved interpretation of the technical parameters is obtained by working with the technical experts in that area. The model equations will also be reviewed and validated as soon as actual cost data is available for DNP-Type projects. Work is in progress to collect cost and technical data from new projects as they enter the implementation stage so that the model may be validated or corrected with improved or actual cost information.

5. ***Linear Estimation Process and Resulting Statistics***

The Ordinary Least Squares (OLS) method was selected to estimate the parameters. OLS is usually recommended when nothing is known about the measurement errors [Beck and Arnold, 1977, § 6.2], since even with little or no information on the error distribution, an adequate predictor may be obtained. However, when information regarding the statistics of the errors is known or assumed, the process produces an efficient estimator of the coefficients (β_j.) This section analyzes the statistical results of the use of this method and identifies the general form of the predictive equation which is the basis for the Cost Estimation Relationships (CER's) which are discussed in the next section.

In order to be succinct in expressing the logic of the process, we will resort to matrix notation in describing the analysis [Beck and Arnold, 1977, § 6.2]. The sum of squares function used for ordinary least squares with the linear model $\eta = \mathbf{X}\underline{\beta}$ is

$$S = (\mathbf{Y} - \mathbf{X}\underline{\beta})^T (\mathbf{Y} - \mathbf{X}\underline{\beta}) \tag{5.1}$$

where $\mathbf{Y}$ and $\mathbf{X}$ are defined by (4.4) and (4.5) respectively and $\underline{\beta}$ is a vector of the undetermined coefficients βj, where, j= 0,n.

Assume that **b** is the estimate of $\underline{\beta}$. Then, since $\mathbf{Y}$ is the estimate of η that is sought,

$$\mathbf{Y} = \mathbf{X}\,b \tag{5.2}$$

In order to solve for the estimated coefficients, **b**, it is necessary to pre-multiply by $\mathbf{X}^T$ so that

$$\mathbf{X}^T\mathbf{Y} = \mathbf{X}^T\mathbf{X}\,\mathbf{b} \tag{5.3}$$

Further pre-multiplication by $(\mathbf{X}^T\mathbf{X})^{-1}$, yields

$$(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y} = (\mathbf{X}^T\mathbf{X})^{-1}(\mathbf{X}^T\mathbf{X})\,\mathbf{b} = \mathbf{b} \tag{5.4}$$

This results in the estimator of the coefficients $\beta_{j'}$

$$\mathbf{b} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y} \tag{5.5}$$

It can readily be demonstrated that 5.5 minimizes the sum of squares, 5.1, and is the OLS estimator of the coefficients βj [Johnson and Wichern, 1988, § 7.3].

For unique estimation of all the coefficients, βj, the matrix $(\mathbf{X}^T\mathbf{X})$ must be non-singular. This means that any one column in $\mathbf{X}$ cannot be proportional to any other column or any linear combination of columns because if such a proportionality exists the determinant of $(\mathbf{X}^T\mathbf{X})$ must equal zero.

As we have mentioned before, if the errors are additive, of zero mean in $\mathbf{Y}$ and $\mathbf{X}$, and the $\underline{\beta}$ are nonstochastic, then E($\mathbf{b}$) is an unbiased estimator of $\underline{\beta}$ such that,

$$E(\mathbf{b}) = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{X}\,\underline{\beta} = \underline{\beta} \tag{5.6}$$

The covariance matrix for the coefficients is expressed as:

$$\mathrm{cov}(\mathbf{b}) = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\psi\ \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1} \tag{5.7}$$

where, $\psi = (\varepsilon\,\varepsilon^T) = \sigma^2\,\mathbf{I}$.

If it is further assumed that the errors are uncorrelated and of constant variance. Then the covariance matrix for the coefficients may be reduced to the expression ,

$$\mathrm{cov}(\mathbf{b}) = (\mathbf{X}^T\mathbf{X})^{-1}\sigma^2 \tag{5.8}$$

which is the minimum covariance matrix of $\mathbf{b}$. The variances of the bj (or the SE^2, depending on the assumptions being used) may be obtained from the diagonal elements of this matrix [Draper and Smith, 1966, § 4.2].

Similarly, from the relationships, 5.5, 5.6, and 5.7, all of the necessary items required to evaluate the fit are obtained. The following table lists the basic data items needed for the analysis:

Table 4. Results of Linear Estimation Process Required for Assessment

	Estimated Coefficients (b_j) and related statistics							
	b_k	b_{k-1}	•	•	•	b_2	b_1	b_0
Est. Value (b_j)	...	...	...	...	...	...	...	...
Std. Error (SE)	...	...	...	...	...	...	...	...
t Statistic	...	...	...	...	...	...	...	...

Statistics on the Estimate (Y)					
R^2	SE (Y)	F	df	SS_{reg}	SS_{resid}
...	...	...	...	...	...

The predicted b_j values and the standard errors for the coefficients are, of course, produced as a direct result of the least squares minimization. For the assumptions on the error being used, the following statements hold,

$$E\,(b_j) = \beta_j \tag{5.10}$$

$$V(b_j) = c_{jj}\,\sigma^2 \tag{5.11}$$

where the cjj are the diagonal elements of $(\mathbf{X}^T\mathbf{X})^{-1}$.
If σ is not known or normality is suspect then

$$\text{est. var}\,(b_j) = c_{jj}\,s^2 \tag{5.12}$$

and,

$$SE(b_j) = (\,c_{jj}\,s^2)^{1/2} \tag{5.13}$$

where s^2 is the sample variance for each bj and $SE(b_j)$ is the standard error of estimate.

In a similar manner the variance of Y, V(Y), or the standard error of Y, SE(Y), can be determined from the diagonal elements of,

$$\text{cov}(Y') = \mathbf{X}\,(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\sigma^2 \tag{5.14}$$

Under the assumptions being invoked, the t statistic for each bj may be computed as,

$$t_j = E(b_j) \;/\; SE(b_j) \tag{5.15}$$

The Coefficient of Multiple Determination, R^2, is defined as,

$$R^2 = SS_{reg} \div SS_{tot} = \Sigma(Y'_i - \underline{Y})^2 \div \Sigma(Yi - \underline{Y})^2 \tag{5.16}$$

where, SS_{reg} is the regression sum of squares (the deviation between the regression line (Y'_i) and the mean ($\underline{Y}$) and SS_{tot} is the total sum of squares (the total deviation between the data (Yi) and the mean ($\underline{Y}$). However, since SS_{tot} is the sum of SS_{reg} and SS_{resid}, the R^2 statistic may be calculated as,

$$R^2 = SS_{reg} \div (SS_{reg} + SS_{resid}) \tag{5.17}$$

where, SS_{resid} is defined as $\Sigma(Y'_i - Y_i)^2$

The F statistic, used in the test for lack of fit is computed as,

$$F(df) = [SS_{reg} \div k] \div SE(Y)^2 \tag{5.18}$$

The F statistic for the fit is dependent on the degrees of freedom, df, which is defined for the table above, as: the number of data points, n, less the number of variables being determined in the regression analysis, k (including the constant, b_0).
The F-test criteria for goodness of fit used is that,

$$F > F_{crit} \tag{5.19}$$

where F_{crit} is the F(k, df, α) critical value from the F-tables. The greater F is than the F_{crit} value, the better the confidence that the "best" fit has been achieved.

6. *Cost Estimation Relationships and Constraints*

The cost estimation relationships, which are the direct expression of the model, are built utilizing the predictive equation (4.6), the coefficients determined in the linear estimation process, and the corresponding statistics described in section 5. This section summarizes the CER's developed for the Spacecraft Subsystem Model by subsystem, including the constraints imposed by the data sets used in the linear estimation process.

6.1. *Attitude Determination and Control (ADCS)*

The following CER for the estimated subsystem cost (Y) in millions of dollars (FY97) was determined for ADCS subsystems within the range of the data domain:

$$Y = b_0 + b_1{}^*X_1 + b_2{}^* X_2 + b3{}^*X_3 \tag{6.1}$$

Coefficients & Constraints for ADCS

X	Parameters	units	Constraints: Avg	S.Dev	Max	Min
X 0	Constant =1	n/a	n/a	n/a	n/a	n/a
X 1	Subsystem Mass	kg	16.06	13.6	48.1	1.9
X 2	D/L Data Rate	kbps	60.63	149	600	0
X 3	Pointing Knowledge	arcsecs	327	302	900	5

Coeff. Symbol	Coeff Value
b_0	9.674
b_1	0.2428
b_2	0.0064
b_3	-0.004

6.2. *Command and Data Handling (C&DH)*

The following CER for the estimated subsystem cost (Y) in millions of dollars (FY97) was determined for C&DH subsystems within the range of the data domain:

$$Y = b_0 + b_1{}^*X_1 + b_2{}^* X_2 \tag{6.2}$$

Coefficients & Constraints for C&DH

X	Parameters	units	Constraints: Avg	S.Dev	Max	Min
X 0	Constant =1	n/a	n/a	n/a	n/a	n/a
X 1	D/L Data Rate	kbps	71.8	1.55	600	0
X 2	Redundancy	ordinal	2.6	0.7	3	1

Coeff. Symbol	Coeff Value
b_0	0.3078
b_1	0.0163
b_2	2.4886

This CER covers the sum of both hardware and software for the C&DH subsystem.

6.3. *Telecommunications (Telecom)*

The following CER for the estimated subsystem cost (Y) in millions of dollars (FY97) was determined for Telecommunications subsystems within the range of the data domain:

$$Y = b_0 + b_1{}^*X_1 + b_2{}^* X_2 + b_3{}^* X_3 + b_4{}^* X_4 \tag{6.3}$$

Coefficients & Constraints for Telecom

X	Parameters	units	Constraints Avg	S.Dev	Max	Min
X 0	Constant =1	n/a	n/a	n/a	n/a	n/a
X 1	Subsystem Mass	kg	14	6	30	7
X 2	Redundancy	ordinal	2.3	0.8	3	1
X 3	X/Ka Band	ordinal	0.5	0.5	1	0
X 4	S/UHF Band	ordinal	0.4	0.5	1	0

Coeff. Symbol	Coeff Value
b_0	10.4
b_1	0.16946
b_2	0.9755
b_3	-3.54
b_4	-6.7623

6.4. *Power Generation (Power)*

The following CER for the estimated subsystem cost (Y) in millions of dollars (FY97) was determined for Power subsystems within the range of the data domain:

$$Y = b_0 + b_1 {}^* X_1 + b_2 {}^* X_2 + b_3 {}^* X_3 + b_4 {}^* X_4 + b_5 {}^* X_5 \qquad (6.4)$$

Coefficients & Constraints for Power

X	Parameters	units	Constraints Avg	S.Dev	Max	Min
X 0	Constant =1	n/a	n/a	n/a	n/a	n/a
X 1	Rad. Dosage	krads	349	972	4000	5
X 2	AMTEC	watts	21.9	58.5	200	0
X 3	Adv Si	watts	2038	3198	10500	0
X 4	GsAs/HT	watts	304	1110	4600	0
X 5	GaAs	watts	553	1900	7900	0

Coeff. Symbol	Coeff Value
b_0	5.08
b_1	0.002
b_2	0.1579
b_3	0.001
b_4	0.002
b_5	0.0022

6.5. *Propulsion*

The following CER for the estimated subsystem cost (Y) in millions of dollars (FY97) was determined for Propulsion subsystems within the range of the data domain:

$$Y = b_0 + b_1 {}^* X_1 + b_2 {}^* X_2 \qquad (6.5)$$

Coefficients & Constraints for Propulsion

X Parameters		units	Constraints Avg	S.Dev	Max	Min
X 0	Constant =1	n/a	n/a	n/a	n/a	n/a
X 1	Ln (S/S Mass)	kg	72.9	59	220.1	7.4
X 2	Ln ISP	n/a	6.1	1	8.2	5.4

Coeff. Symbol	Coeff Value
b_0	-19.7
b_1	3.018
b_2	3.09

6.6. *Structures*

The following CER for the estimated subsystem cost (Y) in millions of dollars (FY97) was determined for Structures subsystems within the range of the data domain:

$$Y = b_0 + b_1{}^*X_1 + b_2{}^*X_2 \qquad (6.6)$$

where Y = Ln (cost).

Coefficients & Constraints for Structures

X Parameters		units	Constraints Avg	S.Dev	Max	Min
X 0	Constant =1	n/a	n/a	n/a	n/a	n/a
X 1	Ln SS Mass	n/a	5	1	6	3
X 2	Ln D/L Data Rate	n/a	1	3	6	- 4

Coeff. Symbol	Coeff Value
b_0	0.65276
b_1	0.33002
b_2	0.00464

Cost is obtained from this CER by computing e^Y.

A supplementary estimate of the mechanical build-up that is usually associated with structures. This CER is,

$$Y = b_0 + b_1{}^*X_1 + b_2{}^*X_2 \qquad (6.6a)$$

Coefficients & Constraints for Mechanical Build Up

X Parameters		units	Constraints Avg	S.Dev	Max	Min
X 0	Constant =1	n/a	n/a	n/a	n/a	n/a
X 1	Subsystem Mass	kg	136	71	337	14
X 2	Pointing Knowledge	arcsec	326	302.5	900	5

Coeff. Symbol	Coeff Value
b_0	1.833
b_1	0.01
b_2	-0.0004

6.7. *Thermal Protection*

The following CER for the estimated subsystem cost (Y) in millions of dollars (FY97) was determined for Power subsystems within the range of the data domain:

$$Y = b_0 + b_1 * X_1 + b_2 * X_2 \quad (6.7)$$

Coefficients & Constraints for Thermal

X Parameters		units	Constraints Avg	S.Dev	Max	Min
X0	Constant =1	n/a	n/a	n/a	n/a	n/a
X1	Redundancy	ordinal	0.8	0.4	1	0
X2	Active/Passive	ordinal	0.1	0.3	1	0

Coeff. Symbol	Coeff Value
b_0	1.817
b_1	1.068
b_2	4.255

6.8. *Statistical Summary*

In evaluating each CER the statistics on the b_j coefficients and the estimated response variable, Y were analyzed. The t, statistics were tested to determine if the resulting estimates for the coefficients were significant contributors. This information was used in determining which coefficients to leave in the regression estimate and which to drop out. In general, the final t statistics satisfied the t-test criteria for significance. The R^2 and the F statistic were used to determine the goodness of fit of the resulting predictive equation for Y. The following table summarizes the estimate statistics associated with the CER's listed above.

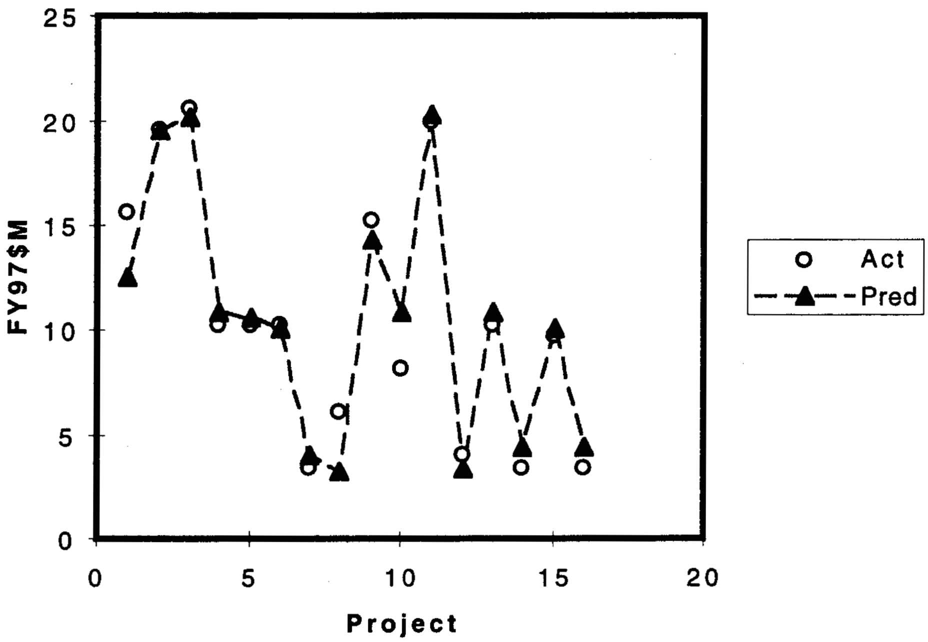
Table 5. Summary Estimate Statistics

Subsystem	R^2	F	k	df	Fcrit	F/Fc
ADCS	.89	33	3	12	5.95	5.53
CDH	.81	24	2	13	6.70	3.55
Telecomm	.88	20	4	11	5.70	3.43
Structures	.76	20	2	13	6.70	3.00
Mech BU	.90	59	2	13	6.70	8.77
Power Gen.	.95	37	2	13	6.70	5.52
Thermal	.74	17	2	12	6.93	2.48
Propulsion	.93	90	2	13	6.70	13.48
Average	**.85**	**29.9**	**2.4**	**12.4**	**6.5**	**4.6**
Min	**.74**	**17.2**	**2.0**	**11.0**	**5.7**	**2.5**
Max	**.9**	**90.3**	**4.0**	**13.0**	**6.9**	**13.5**

From the summary we see that all of the coefficients of multiple determination (R^2) are very high (.74 or above). The F statistics are similarly high and compare well with the F_{crit} values for each of the regression estimates. For this reason, we believe that the estimates produced by the model are accurate predictor's of the Team X estimates for missions that fall within the range of the data base parameters. In order to visually demonstrate how the model is validated against the source data itself, we show (in figure 2) a comparison of actual propulsion subsystem costs (in the data base) with the model predicted costs. Th cost estimate model for this subsystem demonstrates an extremely good fit to the data (R^2 = .95).

This does not mean that all work on the model is complete. Other subsystem models need further refining. A great deal of fine tuning is being conducted as our continuing sessions with the cognizant engineers bring out other causal relations and parameters that need to be validated and tested. It is the goal of the cost team to achieve results such that all of the predictive equations achieve the optimum ability to predict costs within the range of the parameters.

Fig. 2. Propulsion: Actual vs. Predicted



7. *Cost Model Utilization in an Interactive Environment*
The cost model CER's are currently being utilized by Team X in an interactive environment that permits spacecraft designers to see the cost impact of their design decisions as they progress. This permits them to make the necessary trades between, science, technology, and engineering practice to achieve a design that falls within a specific cost cap. Leigh Rosenberg will provide more details of this

process in his paper.

8. *Concluding Remarks*

The Unmanned Spacecraft Subsystem Cost Estimation Model, has evolved into one of the key tools being used to plan and cost advanced missions. The ability to predict what the Team X group of experts would estimate as the cost of a proposed mission is of great value in performing cost trades and off-line studies before calling a Team X session. Besides avoiding unnecessary planning costs, the model permits the cost analyst supporting the Team X sessions to evaluate the costs that are currently being estimated against the model. He may then bring any inconsistencies to the attention of the Team lead and have the issue resolved during the session. In every respect, the model will enhance the efficiency of the planning process and improve the quality of cost estimates for advance projects under study by Team X.

In the future, the model will also be validated against actual project implementation costs as these occur. Once a sufficient number of these new projects have been implemented and the model is modified to reflect these data, the model will become the de facto tool for predicting future project costs which are compliant to the DNP approach.

Current work on the model includes adapting the model to handle non-DNP projects and the addition of a monte carlo simulation feature.

References

Beck, J.V. and Arnold, J.A. , *Parameter Estimation in Engineering and Science,* John Wiley & Sons, Inc., New York, 1977

Draper, N.R. and Smith, H., *Applied Regression Analysis,* John Wiley & Sons, Inc., New York, 1966

Johnson, R. A. and Wichern, D. W., Applied Multivariate Statistical Analysis, Prentice Hall, 1988

Author Biography

George Madrid is a Senior Engineer in the Advanced Mission Operations Section of the Jet Propulsion Laboratory. Mr. Madrid holds a BA Degree in Mathematics from San Francisco State University, a Masters Degree in Systems Engineering from West Coast University, and Certificates in Astrodynamics and Space Communication Systems from UCLA. His technical experience has been in the area of Orbit Determination, Navigation, and Very Long Baseline Interferometry. He has held various engineering and management positions at JPL, the latest being the manager of the Real-Time Operations Element in the Cassini Mission to Saturn. As part of his past experience, he has been involved in both software and hardware cost estimation and is a past member of the International Society for Parametric Estimation (ISPA), Project Management Institute (PMI), and the American Institute for Aeronautics and Astronautics (AIAA).